ABSTRACT OF THE DISCLOSURE

A language model for a language processing system such as a speech recognition system is formed as a function of associated characters, word phrases and context cues. A method and apparatus for generating the training corpus used to train the language model and a system or module using such a language model is disclosed.